



Geiger-Schuller, K., Sforza, K., Yuhas, M., Parmeggiani, F., Baker, D., & Barrick, D. (2018). Extreme stability in de novo-designed repeat arrays is determined by unusually stable short-range interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), 7539-7544. [034102].  
<https://doi.org/10.1073/pnas.1800283115>

Peer reviewed version

Link to published version (if available):  
[10.1073/pnas.1800283115](https://doi.org/10.1073/pnas.1800283115)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via NAS at <http://www.pnas.org/content/early/2018/06/28/1800283115>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

**The extreme stability of de novo designed repeat arrays is determined by short-range interactions.**

Kathryn Geiger-Schuller<sup>1,2\*</sup>, Kevin Sforza<sup>1\*</sup>, Max Yuhas<sup>1,3</sup>, Fabio Parmeggiani<sup>4,5</sup>, David Baker<sup>4</sup>, & Doug Barrick<sup>1§</sup>

<sup>1</sup>Department of Biophysics and Program in Molecular Biophysics, Johns Hopkins University, 3400 N. Charles St, Baltimore MD 21218.

<sup>2</sup>Current address: Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142

<sup>3</sup>Current address: Applied Mathematics, Yale University, New Haven CN 06520

<sup>4</sup>Institute for Protein Design, Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195

<sup>5</sup>Current address: School of Chemistry and School of Biochemistry, University of Bristol. Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 TQ1

\*These authors contributed equally to this work.

§Corresponding Author: barrick@jhu.edu, (410) 516-0409

Classification: Biological Sciences: Biophysics and Computational Biology.

Keywords: Ising model, Repeat proteins, Rosetta, Protein Design, DHRs

## **Abstract**

Designed Helical Repeats (DHRs) are modular helix-loop-helix-loop protein structures that are tandemly repeated to form a superhelical array. Structures combining tandem DHRs demonstrate a wide range of molecular geometries, many of which are not observed nature. Understanding cooperativity of DHR proteins provides insight into the molecular origins of Rosetta-based protein design hyper-stability, and facilitates comparison of energy distributions in artificial and naturally occurring protein folds. Here we use a nearest-neighbor Ising model to quantify the intrinsic and interfacial free energies of four different DHRs. We measure the folding free energies of constructs with varying numbers of internal and terminal capping repeat for four different DHR folds, using guanidine-HCl and glycerol as destabilizing and solubilizing cosolvents. One-dimensional Ising analysis of these series reveals that although inter-repeat coupling energies are within the range seen for naturally-occurring repeat-proteins, the individual repeats of DHR proteins are intrinsically stable. This favorable intrinsic stability, which has not been observed for naturally-occurring repeat proteins, adds to stabilizing interfaces, resulting in extraordinarily high stability. Stable repeats also impart a downhill shape to the energy landscape for DHR folding. These intrinsic stability differences suggest that part of the success of Rosetta-based design results from capturing favorable local interactions.

**Significance statement:**

We apply a statistical thermodynamic formalism to quantify the cooperativity of folding of *de novo*-designed helical repeat proteins (DHRs). This analysis provides a fundamental thermodynamic description of folding for *de novo*-designed proteins and permits comparison to naturally-occurring repeat protein thermodynamics. We find that individual DHR units are intrinsically stable, unlike those of naturally occurring proteins. This observation reveals local (intra-repeat) interactions as a source of high stability in Rosetta-designed proteins, and suggests that different types of DHR repeats may be combined in a single polypeptide chain, expanding the repertoire of folded DHRs for applications such as molecular recognition. Favorable intrinsic stability imparts a downhill shape to the energy landscape, suggesting that DHRs fold fast and through parallel pathways.

\body

## **Introduction**

Linear repeat proteins have proven to be useful model systems in the quest to better understand protein folding thermodynamics. Due to their repetitive primary structures, these proteins fold into linearly extended modular arrays with approximate translational symmetry from repeat to repeat. Unlike globular proteins, where interactions can span across the protein sequence, the interactions of linear repeat proteins are confined to within or between adjacent repeats<sup>1</sup>. This architecture permits the application of nearest-neighbor Ising analysis to extract thermodynamic parameters for folding.

One dimensional Ising analysis has been successfully applied to a number of linear helical repeat proteins<sup>2–5</sup>. This analysis assumes that repeat protein stability can be parsed into intrinsic folding energies of individual repeats and coupling energies at the interfaces between adjacent folded repeats. Previous work characterizing linear repeat proteins derived from naturally-occurring folds shows that individual repeats are unstable. In these proteins, stability (and cooperativity) originates in the favorable interfaces between adjacent repeats.

Owing to their modular architectures, repeat proteins have been used in a number of engineering applications. Consensus ankyrin repeats have been used to select for high affinity binding partners<sup>6–10</sup> and to enhance the activity of engineered cellulases<sup>11</sup>. Repeats from transcription activator-like effector proteins (TALEs) have

been engineered for genome editing<sup>12,13</sup>. Tetratricopeptide repeat (TPR) domains have been fused to molecular chaperones to increase substrate affinity<sup>14</sup>. Expanding to architectures beyond this handful of naturally-occurring linear repeat folds would further enable such protein engineering applications. One promising set of templates is the *de novo* designed helical repeat proteins (DHRs)<sup>15</sup>. This series of constructs comprises a wide variety of native-state architectures that are unrelated to naturally occurring repeat proteins.

Here we characterize the stability of a series of DHR proteins using nearest-neighbor Ising analysis. We find that unlike naturally occurring repeat proteins, both the intrinsic folding and interfacial coupling free energies of DHRs are stabilizing, giving rise to extraordinarily high folding stability while maintaining cooperativity. The favorable local stability of DHR repeats suggests a reduced folding barrier. The observation of favorable local stabilities in DHRs provides insights into the success of current Rosetta-based design, and suggests mechanisms for further DHR-based protein designs.

## **Results**

### **Equilibrium unfolding of Designed Helical Repeat proteins**

To investigate the thermodynamic folding behavior of Rosetta-designed repeat proteins with novel fold geometries, we chose DHR candidates for characterization based on the following criteria: (1) available SAXS and crystal structure data that demonstrate that the target structure is adopted, (2) an absence of cysteine residues to reduce complications associated with disulfide linkages, and (3) experimental evidence

that shows the capped repeat proteins to be monomeric in solution. The proteins DHR9, DHR10.2 (a modified version of DHR10; see below), DHR54, DHR71, and DHR79 (Figure 1A) satisfy these criteria. These constructs have no detectable sequence similarity to naturally occurring proteins (lowest E-values from BLAST search ranging from 0.026 to 4), and span a broad range of sequence (Table S1) and structural features<sup>15</sup>, including both left- and right-handed superhelical architectures. Far-UV CD spectra for four repeat NR<sub>2</sub>C constructs (where N and C represent N- and C-terminal polar capping repeats flanking two internal DHR repeats) for each of these DHRs display characteristic minima at 208 and 222 nm, consistent with folded  $\alpha$ -helical proteins (Figure 1B).

To measure DHR stability, we monitored guanidine-HCl induced unfolding transitions using CD spectroscopy at 222 nm. NR<sub>2</sub>C constructs of DHR10.2, DHR54, DHR71, and DHR79 unfold in a single sigmoidal unfolding transition, which is well-fitted with a two-state model (Figure 1C). DHR9 did not unfold across a range of temperatures, pH, and denaturant concentrations, precluding thermodynamic analysis. The unfolding transitions of DHRs 54, 71, and 79 have high slopes and midpoints for unfolding. The steep guanidine unfolding transitions of these three constructs suggest a high level of cooperativity; two-state fits of the unfolding transitions yield m-values that are similar to predictions from empirical correlation (Table S1)<sup>16</sup>. In contrast, the unfolding transition of DHR10.2 occurs over a broad range of denaturant concentration and has a low midpoint compared to the other DHRs.

## Length and capping dependence of stability

To determine the effects of variation in repeat number and the sequence substitutions associated with the N- and C-terminal capping repeats on stability, we constructed a series of DHR proteins that delete terminal and internal repeats. For some singly-capped constructs, soluble oligomers could be detected by sedimentation velocity analytical ultracentrifugation (SV-AUC). To eliminate oligomerization, glycerol was added to ten percent (volume to volume). SV-AUC demonstrates that in the presence of glycerol, most singly-capped constructs are monomeric (Supplemental Figure 1). For DHR10, deletion of the C-terminal repeat leads to formation of soluble oligomers even in the presence of glycerol. To prevent this oligomerization, we made a series of charged substitutions to solvent-exposed hydrophobic residues in the N-terminal capping repeat (V12K, I14E, V16E, L39R). We refer to this series as DHR10.2. All variants of DHR10.2 are monomeric.

For each of the four DHR series, we measured unfolding curves for constructs with two, three, and four repeats under conditions where constructs remain monomeric. Two repeat constructs contain a single R repeat with either an N-terminal capping repeat (NR) or a C-terminal capping repeat (RC). Three repeat constructs contain one construct with a single R repeat with both N- and C-terminal capping repeats (NRC), or two R repeats with either an N- (NR<sub>2</sub>) or C-terminal capping repeat (R<sub>2</sub>C). The four repeat construct contains two R repeats with both N- and C-terminal capping repeats (NR<sub>2</sub>C). For DHR54 we were also able to purify and characterize a construct containing a single N-terminal capping repeat.



Stabilities of length and capping variants were monitored by guanidine-HCl induced unfolding transitions by CD spectroscopy at 222 nm as described above (Figure 2). For all DHR proteins, unfolding midpoints increase as the number of repeats increases (compare DHR54 N to NR and NR<sub>2</sub>, DHRs 10.2, 71, 79 NR to NR<sub>2</sub>, and all DHRs NRC to NR<sub>2</sub>C). In most cases, when comparing constructs with the same total number repeats, midpoints are lower for constructs with capping repeats than internal “R” repeats (compare DHR10.2 NRC and R<sub>2</sub>C), indicating that capping repeats are generally less stabilizing than internal "R" repeats.

For the DHR10.2 series, adding a C-terminal capping repeat to NR increases the transition slope and midpoint, whereas adding a C-terminal capping repeat to NR<sub>2</sub> increases the slope more than midpoint (compare NR<sub>2</sub> to NR<sub>2</sub>C, Figure 2A). The C-terminal capping repeat gives rise to a larger slope and midpoint than the N-terminal capping repeat (compare NR<sub>2</sub> to R<sub>2</sub>C), suggesting either greater intrinsic stability for the C-cap or a more stabilizing R:C interface.

For DHR54 and DHR71, the unfolding midpoints for N-terminal capped constructs are higher than those for C-terminal capped constructs (compare NR to RC, Figures 2B-C). Whereas for DHR54 capping identity does not affect transition slope, adding a C-terminal capping repeat to DHR71 appears to result in multistate unfolding behavior (compare NR to NRC, and NR<sub>2</sub> to NR<sub>2</sub>C). For DHR79, the N- and C-terminal capped variants are of similar stability (Figure 2D). In general, longer constructs have steeper transitions, although exceptions described above, in which capping repeats unfold prior to the main transition, result in several notable exceptions.

## Ising analysis quantifies intrinsic and interfacial folding free energies for DHRs

Intrinsic and interfacial folding energies were determined using a one-dimensional Ising model. In this model, the conformations of individual repeats are represented as either folded or unfolded. Thus, for an  $n$ -repeat array, there are  $2^n$  configurations represented by the model. The energy of each configuration is determined by the intrinsic folding energy of each repeat ( $\Delta G_i$ ) as well as the coupling ("interfacial") free energies ( $\Delta G_{i-1, i}$ ) between adjacent repeats.

Because the sequences of the N- and C-terminal capping repeats differ from the sequence of central repeats, three intrinsic energies are included in the model ( $\Delta G_N$ ,  $\Delta G_R$ , and  $\Delta G_C$ ). For all DHRs except DHR54, the model includes only one interfacial free energy ( $\Delta G_{i-1, i}$ ). Although it is possible that the free energies between central repeats and capping repeats differ, it is not possible to resolve such differences unless the unfolding energy of the lone cap can be measured. Because we were able to measure an unfolding transition for a lone N-cap repeat for DHR54 (Figure 2B), a separate term for the interfacial energy between a DH54 N-cap repeat and the adjacent R repeat ( $\Delta G_{N, R}$ ) can be fitted.

To account for effects of glycerol on stability, we expanded our standard single-denaturant model to include a linear intrinsic free energy dependence on glycerol. This model was fitted to DHR guanidine-induced unfolding transitions collected at several glycerol concentrations<sup>2,3,17</sup>. By including guanidine HCl unfolding transitions at different glycerol concentrations, we were able to extract the intrinsic ( $\Delta G_i$ ) and interfacial ( $\Delta G_{i,i+1}$ )

free energies in the absence of glycerol. For DHR10.2, DHR54, and DHR79, we assumed that N-cap, central, and C-cap repeats have identical  $m$ -values. For DHR71, fitting required a separate  $m_{\text{Gdn-HCl}}$  for the C-cap repeat.

Figure 2 shows global fits of the Ising model to four sets of DHR unfolding transitions. There are only six shared thermodynamic parameters (free energies and  $m$ -values) for the fits in Figure 2A and 2D and seven shared thermodynamic parameters in Figures 2B and 2C. Global fits also include separate baseline parameters for each unfolding transition. For all DHR series, the data are well-fitted by the Ising model, and result in low and fairly random residuals. The largest non-random residuals are associated with the rather long native baselines associated with some of the longer constructs.

All DHRs have favorable interfacial free energies, similar to interfacial energies seen for naturally occurring repeat-proteins including ankyrin<sup>3,17</sup>, TPR variants (34PR and 42PR arrays<sup>2,4</sup>) and TALE repeats<sup>5</sup>. The intrinsic folding energies of DHRs are also favorable, in contrast with those of naturally occurring repeats. The majority of the capping repeats also have favorable intrinsic stabilities, although they are typically less stabilizing than the internal repeats. The N- and C-terminal caps of DHR10.2 are intrinsically unstable, as is the C-terminal cap of DHR71, consistent with the multi-state transitions seen in panel 2A and 2C. For all DHRs, glycerol is stabilizing, although the effects of glycerol on stability are significantly lower (and somewhat variable among DHR series) than that of guanidine HCl on a molar basis.

## Discussion

By measuring the length and capping dependence on stability of four DHRs families, we have used a 1D-Ising model to quantify intrinsic folding free energies and interfacial coupling free energies. Unlike previously studied helical repeat proteins, which were based on naturally occurring folds, these proteins were generated by *de novo* design. Quantifying the cooperativity of DHRs using the Ising approach provides a new vantage point to compare and contrast natural and designed proteins. The surprising finding that DHRs have intrinsically stable repeats has important implications for understanding the energetic basis for the success in Rosetta design, for the distribution of cooperativity in naturally occurring repeat proteins, and for the shape of the energy landscape.

### **Rosetta algorithms design stable proteins through favorable local interactions**

In the past decade, 1D Ising analysis has been used to dissect folding cooperativity in a variety of naturally-occurring helical repeat protein families<sup>2–5,17</sup>. These proteins have typically been designed using consensus information obtained from multiple sequence alignments, although for some of these series<sup>4,5</sup>, designs were based on genes with nearly identical sequence repeats. Although exact numbers vary, all of these naturally occurring repeat proteins have unfavorable (i.e., positive) intrinsic folding free energies (unfilled red circles, Figure 3A), which are offset by favorable (negative) interfacial free energies (unfilled blue circles, Figure 3B).

The interfacial energies between designed helical repeats are also stabilizing, and span roughly the same range as those of naturally occurring repeat proteins (filled blue circles, Figure 3B). Variation in interfacial free energies among DHRs seems uncorrelated with repeat length, number of interfacial contacts, or surface area buried between repeats (Table S2). However, intrinsic folding energies for DHRs are favorable (Figure 3A), in contrast to all previously measured intrinsic energies for natural repeat proteins<sup>3,17,2,4,5</sup>. This enhancement of intrinsic stability may reflect a fundamental difference between Rosetta-based *de novo* design<sup>15</sup> and natural selection. Based on the findings here, it appears that Rosetta-based design is particularly good at enhancing local stability. Whether this enhancement results from backbone selection in the early stages of design, sequence design in the intermediate stages, or selection for funneled energy landscapes is unclear. We note that the fraction of charged residues in the DHR sequences is significantly higher (with an average of 0.45, Table S1) than the average for all proteins in SWISS-PROT (0.23). An increase in the number of charged residues has been proposed as a mechanism for increased stability in thermophilic proteins<sup>18</sup>, and has recently been seen to correlate with high stability in consensus proteins<sup>19</sup>.

One consequence of the uniquely stabilizing intrinsic folding energies seen for DHRs is a significant enhancement to overall stability. The stability of a tandem repeat array depends on both the intrinsic and interfacial stabilities. The sum of the intrinsic and interfacial free energies gives the stability increment of adding a repeat to an existing folded array (Figure 3C). For naturally occurring repeat proteins, this stability increment derives solely from the interfacial interaction energy and is offset by the

intrinsic energy. For DRH arrays, the favorable intrinsic folding energies add to the interfacial energies, giving rise to an exceptionally large stability increase for adding a repeat to an existing array and resulting in very high native-state stabilities.

**Differences between the energy landscapes of *de novo* designed and naturally occurring helical repeat proteins.** Quantification of the intrinsic and interfacial free energies of repeat proteins using the Ising model allows the energy landscapes of repeat proteins to be represented in meaningful reaction coordinates, scaled using experimentally determined free energies<sup>20,21</sup>. In this representation, the free energies of states where one or more adjacent repeats are folded and paired are plotted as a function of the number of folded repeats and the location of the partly folded structure (N-terminal, C-terminal, or internal; Figure 4). Ignoring lower probability configurations where unfolded repeats are flanked by folded repeats, there are ten configurations in the NR<sub>2</sub>C landscape (Figure 4A).

For ankyrin consensus repeats, which are based on a naturally-occurring repeat family, intrinsic folding energies are unfavorable<sup>3</sup>; thus, all conformations with one folded repeat have high energies, resulting in a large barrier that must be crossed during folding (Figure 4B). Depending on the structure of the transition state for folding, even higher barriers, in which a second repeat is at least partly folded<sup>22</sup> but not yet paired, can further impede folding. In contrast, because the intrinsic folding energies of DHR repeats are favorable, all partly folded configurations are lower in energy than the fully unfolded state under conditions that strongly stabilize folding (Figure 4C, D for

DHR54). Thus, energy landscapes for DHR folding are comparatively smooth and downhill. Moreover, since addition of each folded DHR54 repeat significantly decreases the free energy, the landscape is also very steep, reflecting a strong driving force for folding.

### **Unstable repeats may be a result of natural selection for folding cooperativity.**

In addition to reflecting successful Rosetta design principles, the difference between intrinsic stabilities of natural and designed helical repeats may reflect features imposed by natural selection on natural repeat folds. Instability of local repeats enhances cooperativity, suppressing both the equilibrium formation of partly folded states and the transient formation of partly structured species through a zippering mechanism during folding. Such species may be prone to misfolding and aggregation. Naturally occurring repeat proteins may have evolved to minimize such structures by partitioning stability into long-range versus local interactions. Obviously, there is no such pressure on designed helical repeats. Although many of these species are also suppressed in the unfolding transitions of DHR54 and DHR79 (Figure 2), owing to the strongly destabilizing effects of guanidine the intrinsic stabilities at the transition regions, favorable intrinsic stability would promote conformations where individual repeats are folded relative to the unfolded state. In contrast, for DHR10.2 and DHR71, multistate unfolding is clearly seen for a number of the constructs. This energetic partitioning is consistent with ideas that have emerged from energy landscape theory that natural proteins have been selected to minimize energetic frustration<sup>23–27</sup>. Moreover, family-

specific functional constraints on naturally-occurring repeat proteins may modulate cooperativity to allow for precise conformational fluctuations, as has been suggested for DNA-binding by TALE-repeat proteins<sup>5</sup>.

Lastly, it is possible that nature doesn't select for or against unfavorable intrinsic energies in repeat proteins, but simply selects for global stability above some threshold value<sup>28,29</sup>. Because repeat proteins have very favorable interfacial free energies, global stability can be achieved in combination with modestly destabilizing intrinsic energies. Partitioning stability into interfacial interactions will maintain cooperativity, allowing for functional sequence variation that decreases intrinsic energy. Resolving the intrinsic and interfacial interactions of specific residues will help test these ideas.

## **Methods**

### **Cloning, expression, and purification**

Genes containing DHR repeat constructs were purchased as GeneStrings from GeneArt and cloned with C-terminal His<sub>6</sub> tags via Gibson Assembly. DHR constructs were grown in BL21(T1R) cells at 37°C to an OD of 0.6-0.8, induced with 0.2 mM IPTG, and expressed overnight at 17°C. Following cell pelleting, resuspension, and lysis in 25 mM sodium phosphate (pH 7.0) and 150 mM NaCl, proteins were purified by affinity chromatography on an Ni-NTA column. Proteins were eluted using 250 mM imidazole and dialyzed into 150 mM NaCl, 0-20% glycerol, and 25 mM NaPO<sub>4</sub> pH 7.0.

### **Circular Dichroism (CD) spectroscopy**



Circular Dichroism measurements were collected using an AVIV model 400 CD Spectrometer (Aviv Associates, Lakewood, NJ). Far-UV CD scans were collected at 25°C using an 0.1 cm pathlength quartz cuvette, with protein concentrations of 15-30  $\mu$ M. Buffer scans were recorded and were subtracted from the raw CD data. CD-monitored guanidine unfolding transitions at 222 nm were generated with an automated titrator using 1.5-3  $\mu$ M protein and a 1 cm pathlength quartz cuvette.

### Ising analysis

To determine the intrinsic and interfacial free energies for folding of DHR arrays, and to analyze energies of partly folded states, we used a one-dimensional Ising formalism<sup>30,31</sup>. In this model, intrinsic folding and interfacial interaction between nearest neighbors are represented using equilibrium constants  $\kappa$  and  $\tau$ , respectively, where

$$\kappa_N = e^{-(\Delta G_N - m_{GdnHCl}[GdnHCl] - m_{glycerol}[glycerol])/RT} \quad (1)$$

$$\kappa_R = e^{-(\Delta G_R - m_{GdnHCl}[GdnHCl] - m_{glycerol}[glycerol])/RT} \quad (2)$$

$$\kappa_C = e^{-(\Delta G_C - m_{GdnHCl}[GdnHCl] - m_{glycerol}[glycerol])/RT} \quad (3)$$

$$\tau = e^{-(\Delta G_{i-1, i})/RT} \quad (4)$$

For all DHRs, the intrinsic folding free energies of  $N$  (solubilizing N-terminal cap),  $R$  (consensus repeat), and  $C$  (solubilizing C-terminal cap) are independent adjustable parameters. DHR10.2, DHR71, and DHR79 are well described by a simple model

where the interfacial interactions of the  $N:R$ ,  $R:R$ , and  $R:C$  pairs are identical. DHR54 unfolding transitions are better fitted by a model where the interfacial interactions of the  $R:R$  and  $R:C$  interface are identical, whereas the  $N:R$  pair is different. Glycerol and GdnHCl dependences are built into the intrinsic (but not the interfacial) terms. DHR71 unfolding transitions are better fitted by a model that includes a separate denaturant dependence for the C-terminal cap ( $m_{\text{GdnHCl}, C}$ , Table 1).

Using these equilibrium constants, a partition function  $q$  for an  $n$ -repeat construct can be constructed by multiplying two-by-two transfer matrices:

$$q = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \kappa_N \tau & 1 \\ \kappa_N & 1 \end{bmatrix} \left[ \begin{bmatrix} \kappa_R \tau & 1 \\ \kappa_R & 1 \end{bmatrix} \right]^{n-2} \begin{bmatrix} \kappa_C \tau & 1 \\ \kappa_C & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (2)$$

This representation correlates the each repeat to its neighbor through the separate rows of each matrix. The fraction of folded protein ( $f_{\text{folded}}$ ) can be obtained by differentiation:

$$f_{\text{folded}} = \frac{1}{nq} \left( \kappa_N \frac{\partial q}{\partial \kappa_N} + \kappa_R \frac{\partial q}{\partial \kappa_R} + \kappa_C \frac{\partial q}{\partial \kappa_C} \right) \quad (3)$$

Ising parameters were determined by globally fitting Eq. 3 to guanidine-induced unfolding transitions collected at 0, 10, and 20% glycerol. Fitting was performed using the nonlinear least squares algorithm of the Imfit package<sup>32</sup> using an in-house python program (written by J. Marold<sup>4</sup> and adapted to include glycerol dependence by K.G.-S.) Confidence intervals (95%) were determined by performing 2000 bootstrap iterations.

## Acknowledgements

The authors thank members of the Barrick and Baker lab for their input on this work, and Jeff Gray for helping to initiate these studies. The authors acknowledge the support of the Center for Molecular Biophysics at Johns Hopkins and Dr. Katherine Tripp for instrumental and technical support. Support to KG-S was provided by NIH training grant T32-GM008403. Support for this project was provided by NIH grant 1R01-GM068462 to Doug Barrick.

## References

1. Kloss, E., Courtemanche, N. & Barrick, D. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch. Biochem. Biophys.* **469**, 83–99 (2008).
2. Kajander, T., Cortajarena, A. L., Main, E. R. G., Mochrie, S. G. J. & Regan, L. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* **127**, 10188–10190 (2005).
3. Aksel, T., Majumdar, A. & Barrick, D. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Struct. Lond. Engl.* **1993** **19**, 349–360 (2011).
4. Marold, J. D., Kavran, J. M., Bowman, G. D. & Barrick, D. A Naturally Occurring Repeat Protein with High Internal Sequence Identity Defines a New Class of TPR-like Proteins. *Struct. Lond. Engl.* **1993** (2015). doi:10.1016/j.str.2015.07.022
5. Geiger-Schuller, K. & Barrick, D. Broken TALEs: Transcription Activator-like Effectors Populate Partly Folded States. *Biophys. J.* **111**, 2395–2403 (2016).
6. Forrer, P., Stumpp, M. T., Binz, H. K. & Plückthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2–6 (2003).
7. Steiner, D., Forrer, P. & Plückthun, A. Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *J. Mol. Biol.* **382**, 1211–1227 (2008).
8. Plückthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).

9. Hansen, S. *et al.* Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity. *Sci. Rep.* **7**, 16292 (2017).
10. Wu, Y. *et al.* Rigidly connected multispecific artificial binders with adjustable geometries. *Sci. Rep.* **7**, 11217 (2017).
11. Cunha, E. S., Hatem, C. L. & Barrick, D. Synergistic enhancement of cellulase pairs linked by consensus ankyrin repeats: Determination of the roles of spacing, orientation, and enzyme identity. *Proteins* **84**, 1043–1054 (2016).
12. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757–761 (2010).
13. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).
14. Cortajarena, A. L., Kajander, T., Pan, W., Cocco, M. J. & Regan, L. Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. *Protein Eng. Des. Sel. PEDS* **17**, 399–409 (2004).
15. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
16. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci. Publ. Protein Soc.* **4**, 2138–2148 (1995).
17. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J. Mol. Biol.* **376**, 241–257 (2008).

18. Haney, P. J. *et al.* Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3578–3583 (1999).
19. Tripp, K. W., Sternke, M., Majumdar, A. & Barrick, D. Creating a Homeodomain with High Stability and DNA Binding Affinity by Sequence Averaging. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.6b11323
20. Mello, C. C. & Barrick, D. An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14102–14107 (2004).
21. Aksel, T. & Barrick, D. Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophys. J.* **107**, 220–232 (2014).
22. Ferreira, D. U., Cho, S. S., Komives, E. A. & Wolynes, P. G. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *J. Mol. Biol.* **354**, 679–692 (2005).
23. Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7524–7528 (1987).
24. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
25. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
26. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).

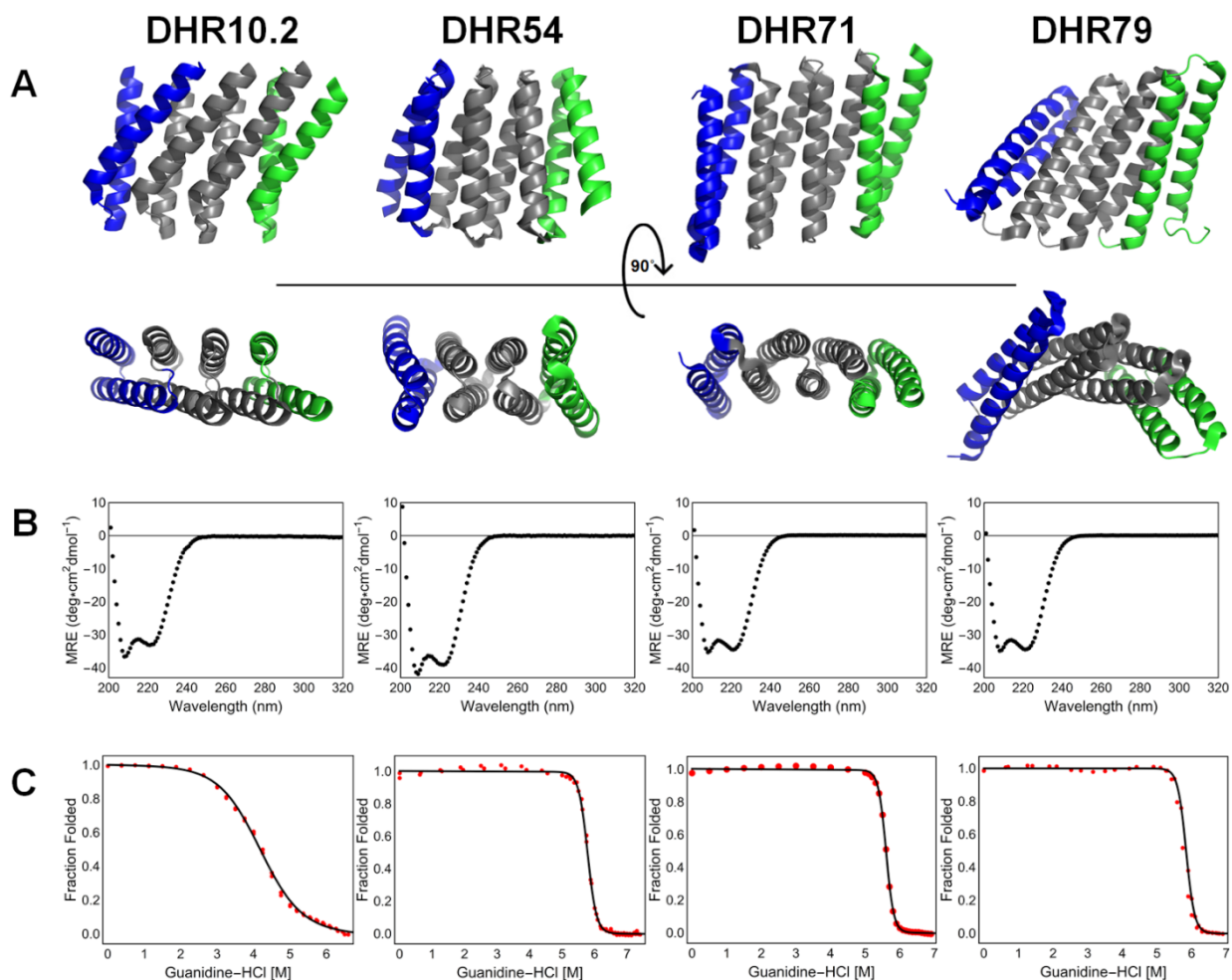
27. Oliveberg, M. & Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **38**, 245–288 (2005).
28. Zeldovich, K. B., Chen, P. & Shakhnovich, E. I. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16152–16157 (2007).
29. Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins* **46**, 105–109 (2002).
30. Aksel, T. & Barrick, D. Analysis of repeat-protein folding using nearest-neighbor statistical mechanical models. *Methods Enzymol.* **455**, 95–125 (2009).
31. Poland, D. & Scheraga, H. A. *Theory of helix-coil transitions in biopolymers: statistical mechanical theory of order-disorder transitions in biological macromolecules*. (Academic Press, 1970).
32. Newville, M., Stensitzki, T., Allen, D. B. & Ingargiola, A. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. (Zenodo, 2014).  
doi:10.5281/zenodo.11813
33. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
34. John Philo's Software Home Page. Available at: <http://www.jphilo.mailway.com/>.  
(Accessed: 6th October 2017)

**Table 1. Thermodynamic parameters obtained from Ising analysis.**

	$\Delta G_N$	$\Delta G_R$	$\Delta G_C$	$\Delta G_{i-1, i}$	$m_{Gdn, i}$	$m_{Glycerol, i}$	$m_{Gdn, C}$	$\Delta G_{N, R}$
DHR10.2	1.46 [1.26, 1.67]	-2.51 [-2.90, -2.15]	0.63 [0.32, 1.00]	-4.80 [-5.10, -4.53]	-1.23 [-1.33, -1.14]	0.36 [0.33, 0.40]	N/A	N/A
DHR54	-0.45 [-0.58, -0.32]	-2.04 [-2.17, -1.92]	-0.84 [-0.94, -0.74]	-6.76 [-6.98, -6.54]	-1.24 [-1.28, -1.21]	0.41 [0.39, 0.43]	N/A	-7.72 [-7.95, -7.49]
DHR71	-3.01 [-3.27, -2.75]	-1.41 [-1.61, -1.23]	3.06 [2.87, 3.29]	-9.93 [-10.50, -9.43]	-1.57 [-1.66, -1.49]	0.17 [0.15, 0.20]	-0.71 [-0.79, -0.64]	N/A
DHR79	-1.84 [-2.06, -1.64]	-3.48 [-3.83, -3.22]	-1.81 [-2.08, -1.61]	-4.83 [-5.14, -4.55]	-1.12 [-1.18, -1.06]	0.15 [0.12, 0.18]	N/A	N/A

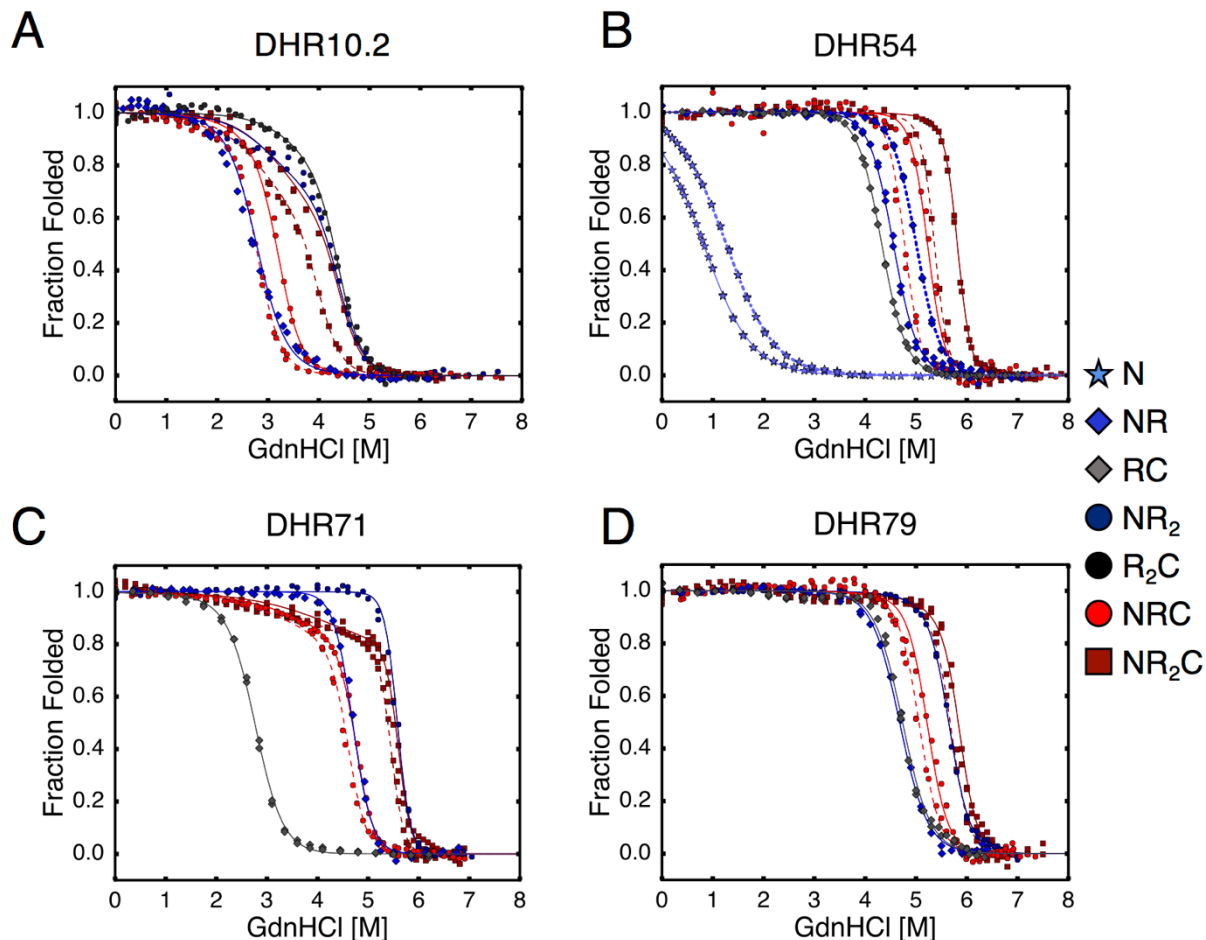
Free energies have units of kcal/mol.  $m_{Gdl}$  and  $m_{Glycerol}$  have units of kcal/mol/[M GdnHCl] and kcal/mol/[M Glycerol]. 95% confidence intervals shown in brackets are from bootstrap analysis with 2,000 iterations.



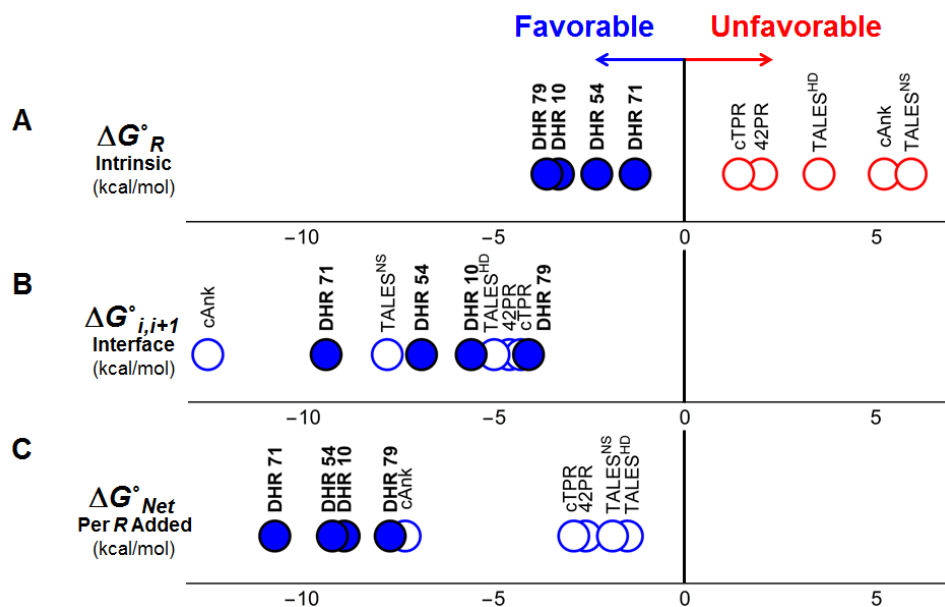


**Figure 1. Structures and stabilities of designed helical repeat proteins.** (A)

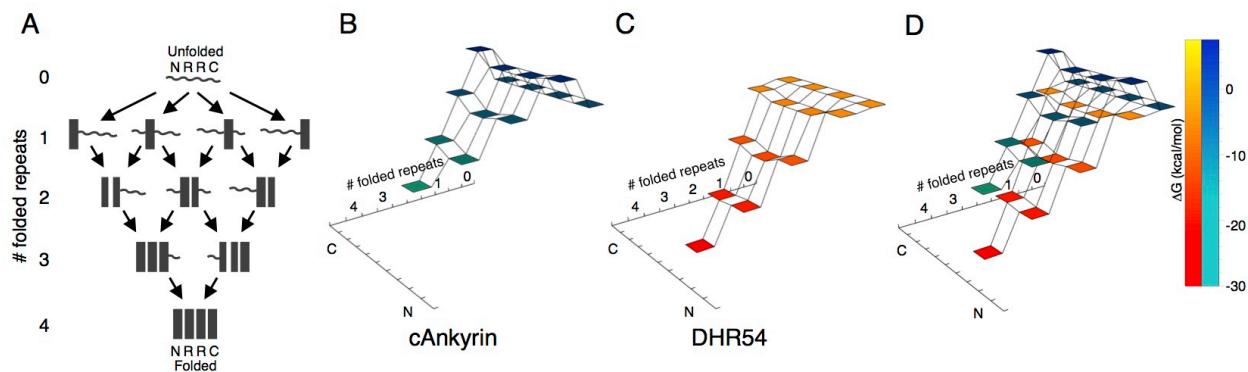
Selected DHR proteins have distinct structures not seen in natural repeat proteins, including unique inter-repeat twists and radii of curvature between repeating units. (B) Far-UV circular dichroism shows characteristic  $\alpha$ -helical spectra for DHR proteins. (C) Guanidine-induced denaturation of four-repeat NR<sub>2</sub>C DHR proteins fit with a two-state unfolding model (black curves) show stable, cooperative folding. Panels in (B) and (C) correspond to the DHR proteins shown in (A). PDB codes are 5CWG (DHR10), 5CWL (DHR54), 5CWN (DHR71), and 5CWP (DHR79).



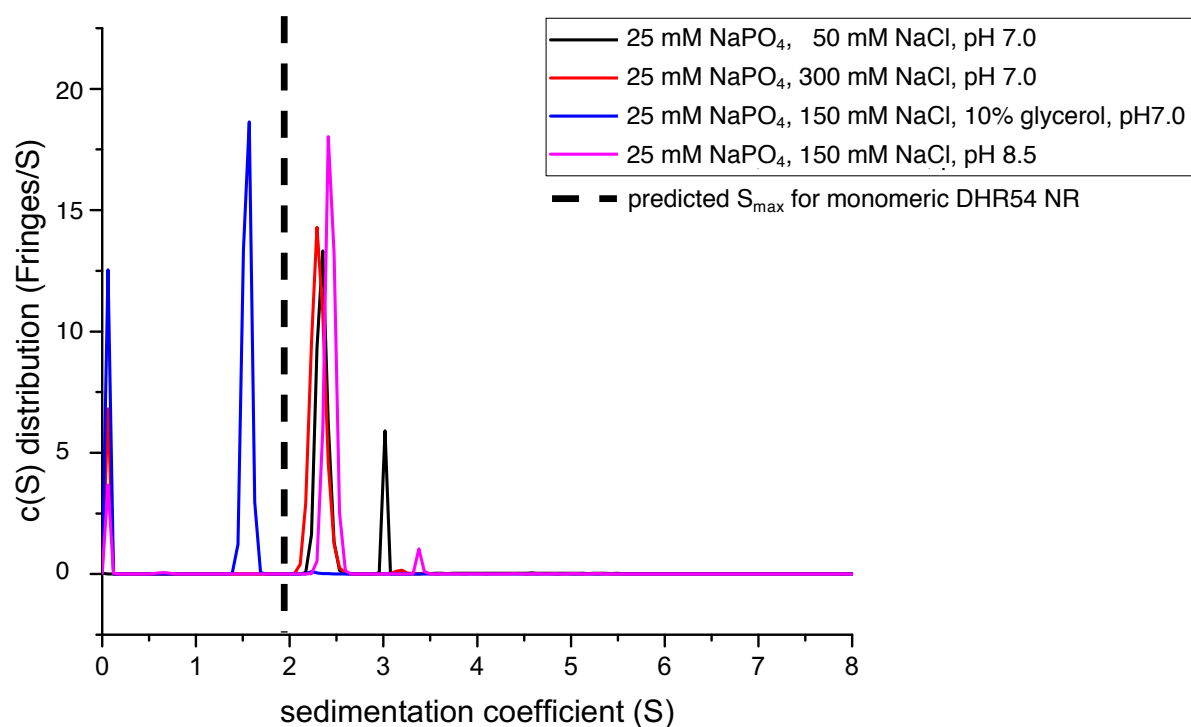
**Figure 2. Unfolding transitions and nearest-neighbor Ising analysis of DHR proteins of different length and capping architecture.** Guanidine-induced unfolding transitions were fitted with a nearest-neighbor Ising model (curves). N-capped constructs are shown in blue, C-capped constructs are shown in grey, and doubly-capped constructs are shown in red. Glycerol concentrations are 0% (dash-dotted curves), 10% (solid curves), and 20% (dashed curves). For all constructs, increasing the number of repeats increases stability (based on unfolding midpoints). Conditions: 25 mM NaPO<sub>4</sub>, 150 mM NaCl, 25°C.



**Figure 3. DHR repeats are intrinsically stable, unlike the repeats of naturally occurring repeat proteins.** (A) Intrinsic folding and (B) interfacial coupling free energies determined by Ising analysis for designed helical repeat proteins (filled circles) and natural repeat proteins (open circles, TALES<sup>NS</sup> and TALES<sup>HD</sup> <sup>5</sup>, 42PR <sup>4</sup>, cANK <sup>3</sup>, cTPR <sup>4</sup>). Unfavorable (i.e., positive) free energy terms are in red, favorable (i.e., negative) free energies are in blue. Designed helical repeats are stabilized by both favorable intrinsic folding and interfacial coupling free energies, whereas natural repeat proteins are destabilized by unfavorable intrinsic folding free energies, which are compensated by large favorable interfacial interactions. (C) Free energy associated with adding a single repeat to a folded array (the sum of intrinsic and interfacial free energies in A and B). Due to their favorable intrinsic folding free energies, DHR proteins are more strongly stabilized by the addition of repeats than natural repeat proteins, resulting in very high overall stability.



**Figure 4. Stabilizing intrinsic energies diminish the barriers on folding energy landscapes for DHR proteins in the absence of denaturant.** (A) Repeat proteins with  $\text{NR}_2\text{C}$  repeat sequences can fold along many pathways. (B-D) Free energy landscapes from experimentally determined intrinsic and interfacial free energies. The vertical dimension (and shading) shows the free energies of partly folded states along the folding pathway shown in (A). (B) Consensus ankyrin repeat proteins, which are based on the naturally occurring ankyrin repeat family, have destabilizing intrinsic energies, and as a result, folding the first repeat results in an early barrier. (C) DHR54 proteins have stabilizing intrinsic folding energies, and as a result, lack this early barrier. Moreover, folding of subsequent repeats is strongly downhill. (D) Overlay of consensus ankyrin (blue-green) and DHR54 (orange-red) free energy landscapes. Landscapes were generated with Mathematica.



**Supplemental Figure 1. Sedimentation Velocity  $c(S)$  plot for DHR54 NR in the absence and presence of glycerol.** Data were processed and fitted in Sedfit<sup>33</sup> as previously described<sup>4</sup>. The predicted  $S_{\max}$  (dotted vertical line) was calculated for DHR54 NR using Sednterp<sup>34</sup>. In the presence of 10 % glycerol, the  $c(S)$  distributions are consistent with monomers.